Analysis of pattern recognition techniques for in-air signature biometrics

Gonzalo Bailador *, Carmen Sanchez-Avila, Javier Guerra-Casanova, Alberto de Santos Sierra

Centro de Domótica Integral (CeDInt-UPM) Universidad Politécnica de Madrid, Campus de Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain

Keywords:
Biometrics
Signature
Accelerometer
Hidden Markov models
Bayes
Dynamic time warping

A B S T R A C T

As a result of advances in mobile technology, new services which benefit from the ubiquity of these devices are appearing. Some of these services require the identification of the subject since they may access private user information. In this paper, we propose to identify each user by drawing his/her handwritten signature in the air (in-air signature). In order to assess the feasibility of an in-air signature as a biometric feature, we have analysed the performance of several well-known pattern recognition techniques—Hidden Markov Models, Bayes classifiers and dynamic time warping—to cope with this problem. Each technique has been tested in the identification of the signatures of 96 individuals. Furthermore, the robustness of each method against spoofing attacks has also been analysed using six impostors who attempted to emulate every signature. The best results in both experiments have been reached by using a technique based on dynamic time warping which carries out the recognition by calculating distances to an average template extracted from several training instances. Finally, a permanence analysis has been carried out in order to assess the stability of in-air signature over time.

1. Introduction

Nowadays, with the advent of smartphones, the original purpose of the mobile phone has been extended to provide many different applications. Nevertheless, this intensive use of mobile devices usually involves the handling of private information whose access must be restricted to an authorized user. In addition to some typical actions such as starting up the device, phoning reserved numbers or reading mail, other new electronic services also require user authentication. For instance, e-commerce, which may take advantage of the mobility provided by these devices, needs to authenticate the user when carrying out electronic transactions [1]. Smartphones have also been proposed to implement a remote control to activate/deactivate the burglar alarm of our homes [2] therefore in this case the user authentication becomes crucial in order to prevent potential intruders. Even though, in places with a low population density, the mobile phone has also been considered for electronic voting [3] to avoid the voters having to visit the polling station which can be far from their houses, so they must also be identified.

In most of these applications, user authentication is carried out by means of a pin code or a password which must be remembered by the user. Nevertheless, in the field of biometrics, this problem is approached by obtaining some features for each individual, which identify him/her unequivocally. Biometric techniques are usually divided into two groups depending on the characteristic used to identify a person, namely physical and behavioural [4]. Physical biometric techniques are based on a physical characteristic preserved in time that a user owns (iris [5], fingerprint [6], hand geometry [7], face [8]) whereas behavioural techniques are related to something that the user is able to repeat in an unique manner (handwriting signature [9], keystroke dynamics [10], gait [11]). In this article, we propose a biometric technique in which a person is authenticated by making his/her handwritten signature in the air (in-air signature) while holding a mobile phone. This biometric technique may be considered not only as a behavioural technique but also as a physical one, since the writing of the signature in the space depends on some physical characteristics of the person (length of the arm, size of the hand holding the device, capability of turning the wrist) [12,13].

In order to capture the movement of the in-air signature, the mobile phone must include a movement sensor. However, this is not a problem since leading mobile phones manufacturers are marketing phones incorporating a 3D accelerometer at a very fast rate. It is expected that in several years, most mobile phones will integrate an accelerometer making this proposed biometric technique accessible for most of the population. For example, Apple sold more than 4 million iPhone mobiles, embedding an accelerometer, just in the first three months of 2009 [14].

Although, there are, to our knowledge, only a few previous works proposing in-air signature as a biometric technique [12,15,13,16] (see Section 2), this problem shares some similarities to other well-studied problems such as gesture and dynamic handwritten signature recognition. In contrast to our problem,

gesture recognition aims to classify some gestures performed by different people to identify the gesture not the person. Nowadays, as a result of the high acceptability of interfaces incorporating an accelerometer, there are many previous works on this topic based on this sensor [17]. In-air signature recognition is also similar to dynamic handwritten signature [18] since both methods represent the signatures by means of temporal signals with spatial information. Nevertheless, in our case the movement is performed in the air instead of on a surface, so it will be much harder for an impostor to copy the 3-D gesture since there are no explicit references of the space where the gesture is performed [19].

In spite of these differences, we consider that the pattern recognition techniques applied to both branches may also be appropriate to solve our problem. This is the case of hidden Markov models (HMMs) and dynamic time warping (DTW) which have been extensively used in these fields obtaining excellent results [20,16]. On the other hand, another commonly used technique consists of the extraction of some global signal features and generating a classifier based on them [21]. Thus, in this article, we have tested all of these pattern recognition techniques in order to decide which is the most suitable one to solve our problem.

To evaluate the feasibility of in-air signature as a biometric technique and the performance of the different classifiers, we have conducted an experiment in which several subjects performed their own signatures in the air. In addition to this experiment, several imitators have attempted to forge previously collected signatures in order to assess to what extent the technique is robust against spoofing attacks. Finally, because in-air signature is a behavioural biometric technique, we have also analysed the stability of in-air signatures over time using the technique which provided the best results. The results for these three experiments will be presented throughout this paper.

The rest of this paper is divided into eight sections. The second section presents the results obtained in previous works in this field. Then, in Section 3, we describe the problem studied in this paper and the evaluation protocol. After that, we explain how all the experiments were conducted and how the signal was captured and normalized. the fifth section describes the application of the different techniques—HMM, Bayes classifiers and DTW—to solve our problem. Afterwards, the results obtained by previous techniques are compared and, in Section 7, we perform the permanence analysis of this biometric feature. Finally, the last section concludes the feasibility of this approach and establishes the most appropriate techniques.

2. Previous works

As we explained before, there are, to our knowledge, only a few previous works proposing in-air signature as a biometric technique [12,15,13,16]. Nevertheless, in these works, the analysed subjects did not use their own handwritten signature but some gestures which were usually less complex than their handwritten signatures. Most of these works [12,15,13] used predefined gestures so the authentication was carried out by measuring the differences among the performances of the same gesture by several individuals. On the other hand, the experiment presented in [16] may be considered closer to our approach as the gestures are different among them. Each gesture was chosen by each subject who participated in the experiments.

In these previous works, two different pattern recognition approaches have been selected to generate the classifiers, which make the authentication of the individuals possible.

The first approach consists of extracting some global features from the acceleration signal like the average or the curvature moments and then training a specific classifier for each individual based on them [12,15]. However, these works are preliminary studies so we consider their results as rather non-conclusive. On the one hand, the article [15] does not provide results on their accuracy since it is only a feasibility study. On the other hand, the authors of [12] reached over 95% recognition rate when discriminating between individuals but only 10 subjects participated in the experiment.

Other works [13,16] have used a template matching approach which carries out the classification by calculating the distances between the given gesture and the template corresponding to each subject. Before calculating these distances, both works use an algorithm based on dynamic programming to align the compared gestures in order to deal with the time variability usually present in these signals.

In more detail, the authors of [13] carried out an experiment in which 12 subjects drew a pentagram several times in the air during 6 weeks. Their algorithm obtained an equal error rate (EER) of 14.7% when attempting to discriminate between individuals in basis to the different ways of making the same gesture. This algorithm compared the similarity of the given gesture to the first gesture, which was considered as a template. The authors realised that the reason for this high rate was that the drawing of the pentagram for each subject had slightly changed over the period. Therefore, they proposed an algorithm to update this initial template achieving an EER of 4%.

In article [16], the authors conducted an experiment over 10 subjects in which each subject chose his/her own gesture. This work reached an EER of 3% by using dynamic time warping to distinguish the gestures of different individuals. However, the authors also carried out an experiment consisting of a spoofing attack in which the rate increased to 10%. Therefore, this result shows that this technique may be weak against this type of attack. Finally, the authors of this paper carried out a survey on the reasons for choosing the gestures concluding that the main ones were the uniqueness of the gesture and ease of remembering. Hence this survey confirmed our decision of using the handwritten signature in the air since it is considered unique and the subjects get used to performing it frequently.

3. Problem statement

The problem presented in this paper consists of authenticating a user by writing his/her signature in the air using a phone which incorporates an accelerometer. As opposed to identification problems in which a given instance is evaluated on the models of different subjects to identify the owner of the analysed signature, in our problem the given instance is only compared to a model of the signature of the genuine user in order to verify his/her identity. Depending on the result of this verification, the mobile system allows or denies access to the user.

In order to extract the pattern of a user's signature, he/she must enrol in the system by writing his/her signature several times. When comparing a given instance to the user's model, the verification system will produce a score which reflects the similarity between the instance and the pattern. Therefore, to accept or reject a specific signature, we must define a threshold for this score. If the score produced for an analysed signature is higher than the fixed threshold then this signature is accepted as belonging to the subject, otherwise it is rejected.

In the evaluation of the performance of this identity verifier we must consider two types of error: either the genuine user attempts to access the mobile system but he/she is rejected (false rejection rate (FRR)) or an impostor is able to enter the system by performing a gesture similar to that of the genuine user

(false acceptance rate (FAR)). The first type of error may be evaluated by checking the system with some testing samples different from the ones used during the enrolment phase. For the second type of error, we have proposed two different attacks. First, a zero-effort impostor attack which measures the similarity between signatures of different subjects by attempting to access the system with the signature of another subject. Second, a spoofing attack is carried out to evaluate the security of the system against impostor's attacks. In this active attack, the impostor attempts to emulate the signature of the user after seeing his/her performing the signature.

Therefore, for both attacks we must find the threshold which provides the best trade-off between FAR and FRR. Some applications may prefer a low FAR instead of a low FRR or vice versa, but in this paper we assume that the best trade-off is obtained when FAR and FRR reach the same value which is called equal error rate (EER). The EER for a zero-effort impostor attack and the EER for active impostor attack will be used throughout this paper to evaluate the different pattern recognition techniques. Receiver operating characteristic (ROC) curves will also be presented for each technique in order to show the relationship between FAR and FRR because they are commonly used in the evaluation of pattern recognition techniques. Furthermore, in this ROC curve we have plotted a diagonal line which represents the points in which FAR and FRR are equal, so the points where the ROC curve cross this line will represent the EER.

Finally, after deciding the most appropriate technique, we have also carried out a permanence analysis to assess the stability of this biometric feature over time. Because an in-air signature is a behavioural characteristic, it may suffer slight variations between performances caused by a different way of holding the phone, changes in clothing, etc.

4. Experimental setup

For the evaluation of the pattern recognition techniques, we have captured the in-air signature of 96 subjects (68 males and 28 females) with an age ranging from 18 to 60 years old. They were asked to repeat their handwritten signature in the air eight times using a device with an embedded accelerometer. During this acquisition process, all of the individuals were recorded on video while performing the signature. Therefore, we have captured 768 genuine signatures (96 subjects $\times 8$ repetitions). However, as a result of some technical problems a 1.95% of the available gestures were corrupted, so we removed these erroneous signatures from our database in order to maintain its validity (753 valid instances).

The active impostor attack has been carried out by six different individuals (four males and two females, with an age ranging from 22 to 29 years old). Each impostor attempted to emulate the signature of all the subjects of the database seven times. These attempts were divided into two sessions: one simulating that the impostor saw the signature by chance and other simulating that the impostor recorded a video with the signature and studied it carefully. In the former session, the impostor only watched the recorded signature twice and then attempted to repeat the signature three times. In the latter session, the impostor could watch the recorded signature as many times as he wanted and then repeat the signature four times. Furthermore, in order to help the impostors to reproduce the signature, they were informed of the name of the individuals, as in most cases, it appears on the signature. Nevertheless, in the results shown in this article we do not distinguish between both results since the differences were minimal. Therefore, for each impostor, we will have 672 (96 subjects $\times 7$ attempts) forgeries of the signatures.

For the permanence analysis, the temporal evolution of a subset of the in-air signatures from the first experiment has been studied. Specifically, eight subjects have repeated their signature in 20 separated sessions over a period of approximately two months. In each session, the subject repeated his/her signature five times. The time interval between consecutive sessions was about five days and each session was scheduled at a different time. Furthermore, there were no restrictions on the clothing of the subjects.

4.1. Data capturing

All signatures have been captured using the mobile device "iPhone" from Apple Inc., which incorporates a tri-axial accelerometer. The three acceleration signals (A_x,A_y,A_z) are measured in gravity units (g) in the range of $[-2.3g, 2.3g]$. This vector is captured with a sampling rate of 100 Hz, which is fast enough for our purpose since the maximum frequency of hand gestures is about 10 Hz [22]. The segmentation of the signatures was performed manually. Before the signature was written, the subject pushed a button on the screen and he/she pushed this button again at the end of the signature. The average length of the in-air gestures performed by all individuals was $4.17\,\text{s} \pm 1.2$ with a maximum duration of 6 s. Fig. 1 shows an example of the acceleration signals captured during the performance of an in-air signature.

4.2. Preprocessing

Because the signatures may be performed at different speeds, the magnitude of the acceleration signals (A), which represent the signatures, may vary considerably. Furthermore, slight changes in the orientation of the sensor may make the contribution of the gravity modify the offset of the signal. For these reasons, we have normalized the signals (A^N) corresponding to each signature in order to facilitate their comparison. For each axis (x, y, z), we have removed the offset (average (μ)) of the signal and we have normalized its amplitude by dividing the signal by its standard deviation (σ) during the signature:

$$A_x^N = \frac{A_x - \mu_x}{\sigma_x} \quad A_y^N = \frac{A_y - \mu_y}{\sigma_y} \quad A_z^N = \frac{A_z - \mu_z}{\sigma_z}$$

Nevertheless, we have not performed normalization on the time dimension since the pattern recognition techniques analysed in this paper are able to deal with the variability of this dimension. On the one hand, Bayes classifiers are based on statistical

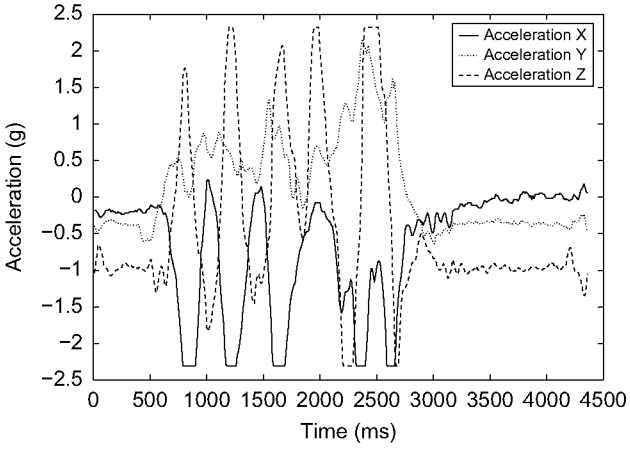

Fig. 1. Example of the acceleration signals of an in-air signature.

features of the data which should be independent of an overall rescaling of the time dimension. On the other hand, DTW is a well-known technique given its ability to align temporal signals. In the case of HMM, some previous works in gesture recognition have proposed to rescale the tempo of the gesture [20] whereas other works did not perform any type of time normalization [27]. Furthermore, the length of in-air signatures captured in our experiment ranges from 1 to 6 s depending on the subject so we consider that this duration is also a characteristic feature of user's signature.

5. Analysed pattern recognition techniques

In order to obtain comparable results for all the following analysed techniques, we have defined some constraints which all methods must follow. First, the training set necessary to learn the model of each signature will be made up of five training instances, which will be chosen randomly from among the eight available ones. Because of this random selection of the training instances, all tests carried out in this paper will be repeated 10 times with different training sets to obtain results independent of the chosen training instances. Therefore, the results presented in this paper will be referred to the average result of these 10 repetitions.

5.1. Hidden Markov models

Hidden Markov model (HMM) is a well-known technique as a result of its ability to model dynamic systems and to deal with noisy data. Initially, HMMs were applied extensively to speech recognition [23], obtaining admirable results. This success made their use widespread in other fields like handwritten character recognition [24] or even activity recognition [25]. Nowadays, HMM may be considered the state-of-art technique in gesture recognition from acceleration data [20,26–28].

HMM consists of an underlying Markov process whose state cannot be directly observed. This hidden state can only be inferred from an output variable that is influenced by this state. There are two different kinds of HMM related to the type of this observable output: discrete and continuous HMMs.

The outputs of discrete HMMs can only take values in a discrete alphabet. Where the pattern to model has continuous outputs, such as the acceleration values of in-air signatures, the output must be discretized. Some works in gesture recognition have used discrete HMMs by partitioning the output space with k-means algorithm [26] or dividing this space into 3D cells of equal volume [28]. However, in this work, we decided to use continuous HMM in order to avoid this discretization process, which sometimes supposes a great loss of signal information.

In contrast to discrete HMMs, the outputs of continuous HMMs can represent any n-dimensional vector with real values, since the output of each state is defined by means of a probability density function on a n-dimensional space. This output function may be modelled with only one Gaussian distribution as in Gaussian HMM (GHMM) or by the combination of several Gaussian functions as in the case of mixtures of Gaussian HMM (MGHMM). Although, most previous works have used GHMM obtaining high recognition rates [20,27], in this paper we have also tested MGHMM since they can provide more complex distributions. Notice that GHMM may be considered as a MGHMM with only one component.

In addition to the type of the output, other HMM parameters must be fixed in advance before training the models: the HMM architecture and the number of states. The HMM architecture refers to the way of connecting the different states and, usually, two architectures are proposed: left–right and ergodic. In left–right architecture, there are only connections between consecutive states and the process can only move forward in these states. On the one hand, this architecture has been successfully used in gesture recognition [27] since it is suitable for temporal signals. On the other hand, ergodic architecture allows any state to be connected to other state without any restriction. Therefore, this provides a powerful model, which is also well suited to gesture recognition [26] since it can express any relation between states. As regards the number of states, previous works have proposed a wide range of values: 5 in [26], 10 in [27] or 12 in [20]. However, the gestures used in these works usually have a duration of less than a second whereas our in-air signatures can take up to 6 s. For this reason, we must consider the inclusion of a greater number of states as well as in other works in which MGHMM was applied to recognize handwritten signatures [29].

Because there were no unified criteria about the type of architecture and the number of states and mixtures used for the outputs, we have carried out several tests to decide which are the best parameters for our application. After some initial analyses, we established that the architecture of the HMM will be ergodic (Fig. 2 details this chosen architecture but only for four states) since it obtained higher recognition rates than left–right one. In order to show how the number of states and mixtures of outputs affects the performance of HMM in our problem, in Fig. 3 we have represented the average EERs of a zero-effort impostor attack obtained for different configurations. In particular, we have tested GHMMs and MGHMMs with mixtures of two and three Gaussian functions and with a number of states ranging from 10 to 70. The lowest EER has been obtained for the configuration with 60 states and three mixtures, so we have selected this configuration for the following experiments. Furthermore, it can be seen that the inclusion of additional states does not improve the results.

The training stage consists of modelling an MGHMM for the in-air signature of each subject. Each MGHMM is learned with

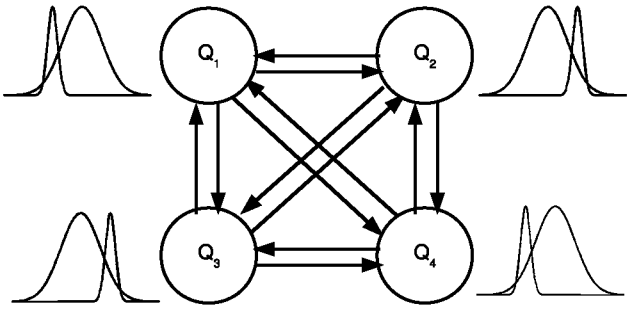

Fig. 2. Example of ergodic GMHMM with four states.

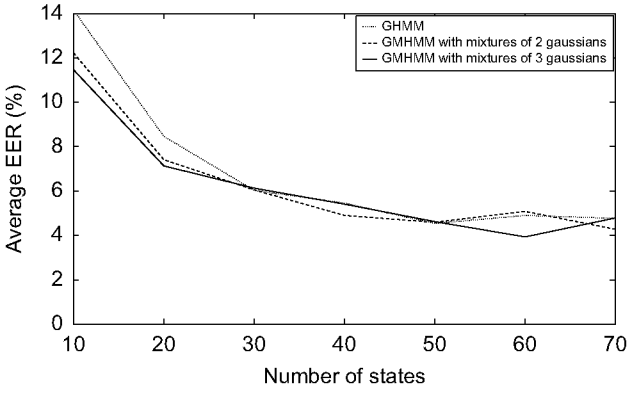

Fig. 3. Average EERs for different combinations of number of states and mixtures of Gaussian functions for the outputs.

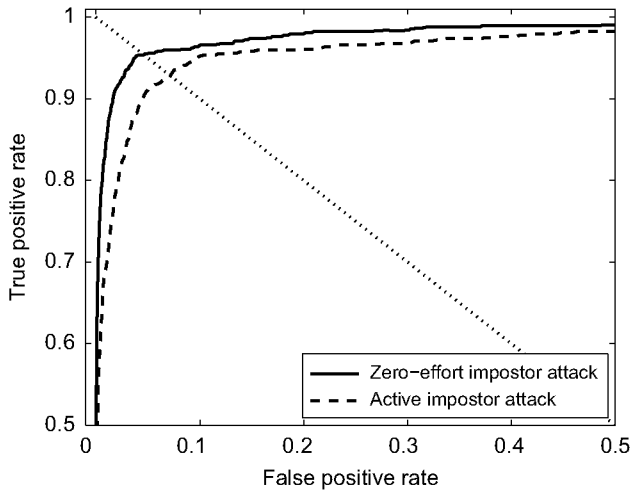

Fig. 4. ROC curves for MGHMM approach trained with five instances.

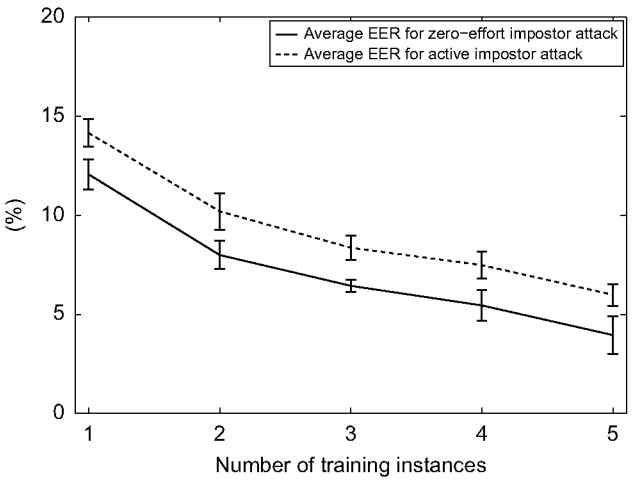

Fig. 5. Performance of HMM for training sets with different sizes.

different instances of the signature by means of an expectation-maximization algorithm. In this paper, this parameter estimator has been iterated 20 times for every trained model using these initial parameters: random transition probabilities and a random output distribution for each state. During the classification process, the signature of each subject is evaluated over the MGHMMs of all subjects. Using the forward algorithm, the probability of each MGHMM to produce this particular signature is computed. Therefore, in this case, the threshold to determine acceptance or rejection of a signature will be fixed on this probability.

In Fig. 4, we have represented the ROC curves obtained by using HMMs with the chosen configuration and learned with five training instances. The solid line shows the results corresponding to the zero-effort impostor attack whereas the dotted line shows the results of the active impostor attack. Although both curves have a similar shape, the ROC curve of the zero-effort attack is above the other ROC curve since the signatures of different subjects are more separable than the signatures from their forgeries. As we explained before, the points where the diagonal line crosses the ROC curves represent the EER. These points have been obtained by placing an overall threshold over the log-likelihood measure produced by each HMM when evaluating a given signature. This measure represents to what extent a signature fits a given HMM and it ranges from 0 when it fits perfectly to $-\infty$ when the signature is completely different to the model. In our case, the threshold values to obtain the EER were -964 for zero-effort attack and -850 for impostor attack which means that it is necessary to be less permissive in order to reject the more impostor signatures.

Previous analysis was carried out by using five training instances, however, in a practical application this will suppose that the subject should repeat his/her signature five times to enrol in the system for the first time. This initial task may be considered annoying for the user, so we have analysed whether this number of training instances may be decreased without reducing considerably the performance. Fig. 5 shows the average EERs for both attacks when using training sets with a size ranging from 1 to 5 instances. Depending on the needs of the application, the user could reduce the number of enrolment instances at the expense of increasing the EER.

5.2. Statistical classifier

In this approach, every signature is described using a set of D features that describe some global characteristics of the signature. These D features can be represented in a space of D dimensions; therefore, each signature can be viewed as a point in this feature space. The goal of a classifier is to obtain the decision boundaries on the feature space from several training instances, so that they allow a given signature to be assigned to the correct class (subject). In particular, for this paper, we have used a Bayesian classifier, which models the class of each signature by means of a multivariate normal density function. In other papers [21] the authors propose to use support vector machines instead of Bayesian ones since this method usually achieves higher recognition rates. Nevertheless, this method requires signatures from other subjects to generate every classifier, whereas Bayesian classifiers can be trained without them. In our case, this is a marked limitation since this system will be developed in a mobile device and therefore the signatures of other users will be not available.

The feature selection is critical to obtain high recognition rates. Statistical features, e.g., mean, standard deviation and the like, are usually used in time series as they describe the global behaviour of the pattern [30]. Features based on frequency like the coefficients of the discrete Fourier transform (DFT) or the discrete time wavelet transform provide information about the different frequencies present in the pattern [31]. Some works have also proposed more complex features like the frequency domain-entropy [30] or Doppler spectrum [32]. In this article, we have used the following features, which were proposed by the author of [21] for gesture recognition using a 3D accelerometer (each feature has been calculated for each axis except for the correlation which is calculated for every pair of axis):

- Mean μ: Average of acceleration values during the signature (μ_x, μ_y and μ_z).
- Standard deviation σ: Standard deviation of acceleration values during the signature (σ_x, σ_y and σ_z).
- Correlation γ: Correlation between the acceleration signals of different axis (γ_{xy}, γ_{yz} and γ_{xz}).
- Entropy δ: Normalized information entropy of the DFT component magnitudes as explained in [21] (δ_x, δ_y and δ_z).
- Energy ε: Sum of all the squared DFT component magnitudes except the DC component of the signal because it has been considered as the independent feature μ.

In order to analyse the best features for discriminating between different signatures, we have generated several classifiers trained with different sets of features. Furthermore, following the indications of the authors of [21], we have divided the signal into several

frames of equal length to calculate these features individually for each frame. This division allows the number of features to be multiplied by the number of frames, which may help the discrimination between subjects. In particular for this article, we have analysed the performance when dividing the signal into 1, 2 and 3 frames. Fig. 6 shows the average EER in a zero-effort impostor attack obtained for the classifiers trained with different combinations of features and number of frames. It can be seen that the best rates are obtained for feature sets which combine the average (μ), the standard deviation (σ) and the correlation (γ). On the other hand, the inclusion of δ and ε features only decreases the performance of the classifier. As regards the number of frames, the figure shows that the division into three frames is only worthwhile when the feature set contains a few number of features, however, in other cases it does not improve the results. We have chosen the feature set that provides the lowest EER which is made up of the average μ, the standard deviation σ and the correlation γ and each sample is divided into three frames. Hence, every signature will be represented with 27 features (three frames $\times$ three features $\times$ three axis).

The ROC curves represented in Fig. 7 show the different behaviours of the classifiers for both attacks. For the zero-effort impostor attack, the classifier produces a high recognition rate for their own signatures and a low acceptance rate for the signatures of other users. This means that the Bayesian classifier generated

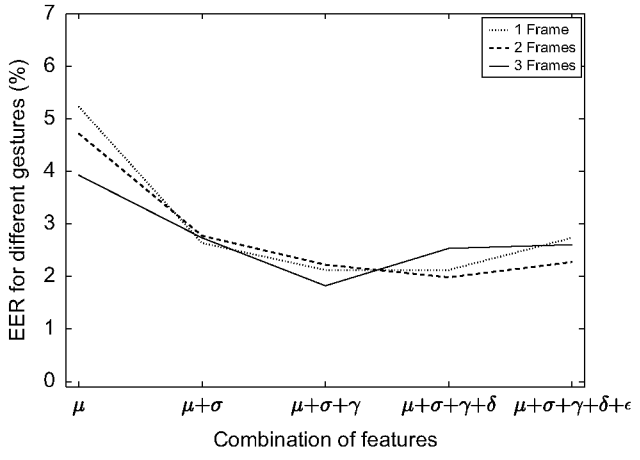

Fig. 6. Choice of best features and number of frames.

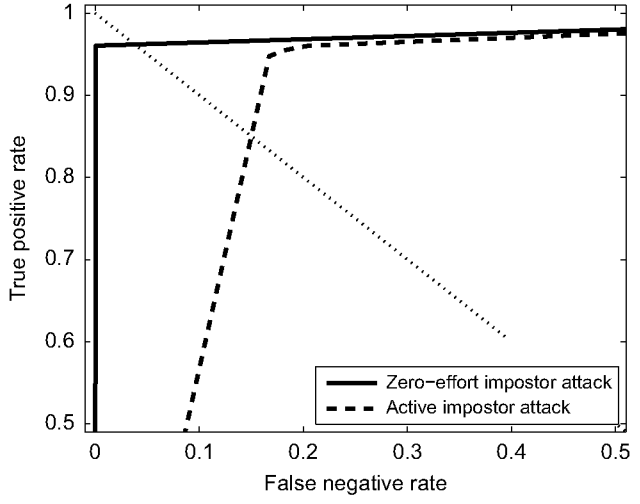

Fig. 7. ROC curves for classifier approach using five training instances.

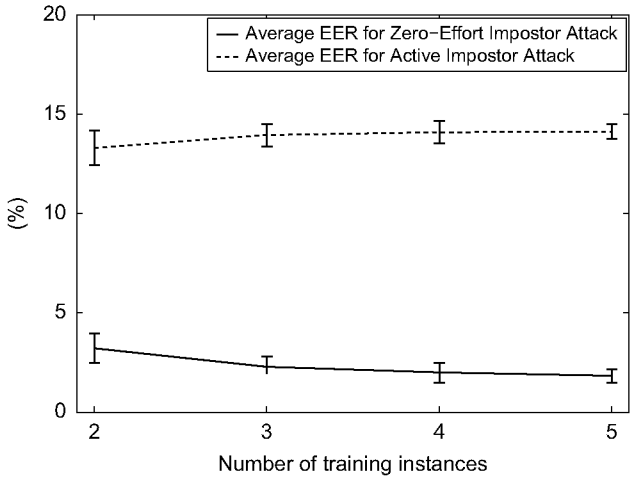

Fig. 8. Performance of Bayesian classifiers for training sets with different sizes.

with the chosen features allow separating the signatures of different subjects. However, the ROC curve for the active impostor attack shows a high acceptance rate for genuine signatures and their forgeries. Therefore, this classifier is not able to discriminate between a genuine signature and a forged one. This fact can be also observed in the threshold values which produce the EERs. In the zero-effort attack, the threshold value is almost 0 which makes it easy to accept any signature of the same subject, whereas in the impostor attack this value is close to 1 in order to reject those signatures performed by an impostor.

Finally, the performance analysis of Bayesian classifiers when using training sets of different sizes is shown in Fig. 8. This analysis has not been carried out for training sets with only one instance as the extraction of the covariances of the multivariate Gaussian distributions requires at least two training instances. In this figure, the EER for zero-effort impostor attack decreases smoothly as the number of training instances increases. For instance, the rate obtained for two training instances is 3.2% and for five training instances is 1.8%. Therefore, this smooth variation makes it possible to fix the enrolment time depending on the needs of the application. Nevertheless, this figure also shows that the EER for the active impostor attack remains stable (at about 14%) in spite of generating the classifiers with an increasing number of training instances.

5.3. *Dynamic time warping*

Template matching approaches are based on the fact that signals belonging to the same pattern have a similar shape. The similarity between two different signals is measured with a determined distance, which is calculated throughout the extension of the signals. Most common distances, as Euclidean distance, can only compare sequences that have the same length. However, in the real world, two signals belonging to the same temporal pattern can present variability not only in their amplitude, but also in their duration or speed. Moreover, this variability can be different in several parts of the signal which might not be corrected with an overall rescaling of the sequences.

DTW has been widely used in many fields such as speech recognition [33], and even gesture recognition [34]. In these fields, this technique has shown a great ability to calculate the similarity between signals that present variability in the time axis. When comparing two signals, this method compresses and expands the time scale of both signals in order to align them and to minimize the total distance. Therefore, because of the nature of the acceleration signals of the captured signatures, this distance

seems appropriate to the task at hand. Similar distances based on DTW has been proposed to deal with acceleration signals as the derivative DTW (DDTW) [35], which aligns the signals in basis to the estimated derivative instead of the signal itself. However, we have tested DDTW using the signatures of our database without obtaining any improvement, so we have used the classic DTW.

In order to make the signal alignment, the DTW algorithm generates a matrix M with the accumulated costs of aligning the samples of two signals (A, B). Each element m_{ij} represents the accumulated cost of aligning subsequences of A and B from the beginning (A_1, B_1) to the elements A_i and B_j. In our case, the cost of aligning to points A_i and A_j is the Euclidean distance between the 3D acceleration vectors of these points. After that, the algorithm determines the path of this matrix which allows both signals to be aligning, thus producing the minimum cost. This path W is called the "warping path" and represents how both signals must be extended or compressed in order to be aligned. Hence, the distance between two signals will be the sum of each element cost of the chosen warping path. This distance will be divided by the length of the warping path W in order to make it independent of the length of the signals. An example of the DTW algorithm for two signatures is represented in Fig. 9 where matrix M is depicted with the accumulated costs and the warping path W using a white line.

The normalized histograms represented in Fig. 10 show the distribution of the distances between signatures of the same subject, signatures of different subjects and their forgeries. The intersection of both distributions represents those signatures that may be misclassified. It can be seen that the intersection between the distributions of distances between the same signatures and the distances between different signatures is insignificant and

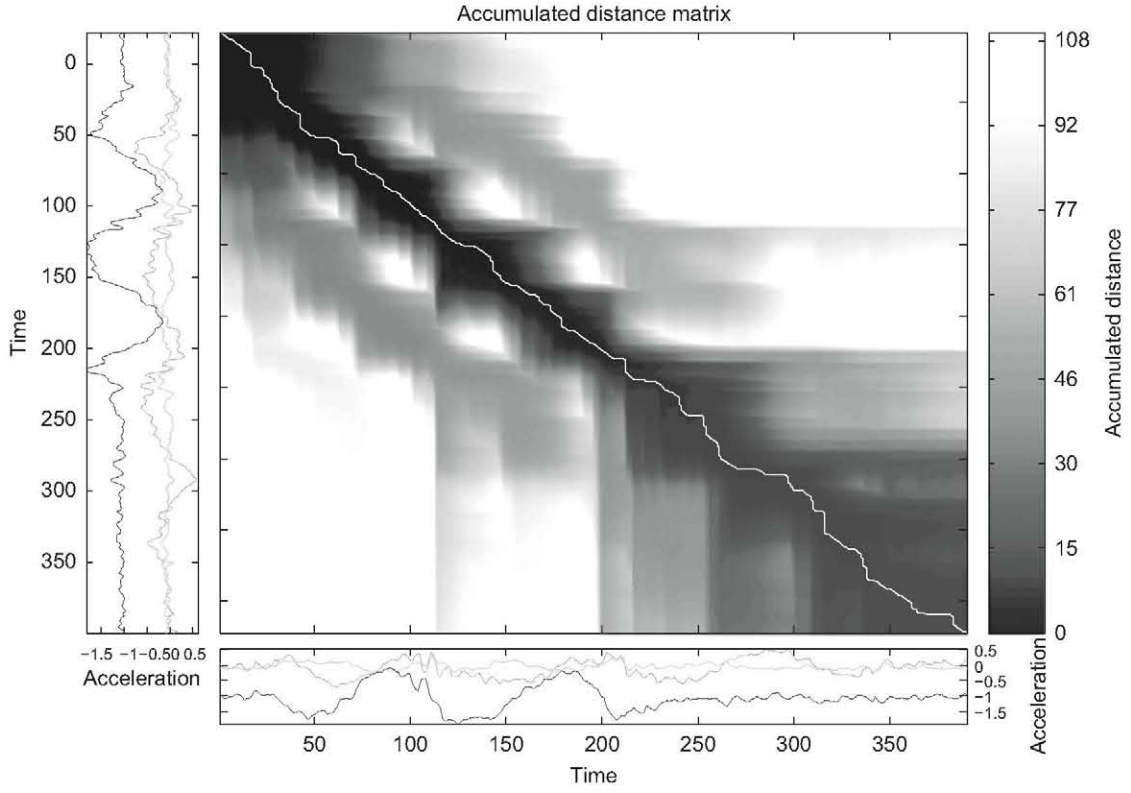

Fig. 9. Example of alignment using DTW.

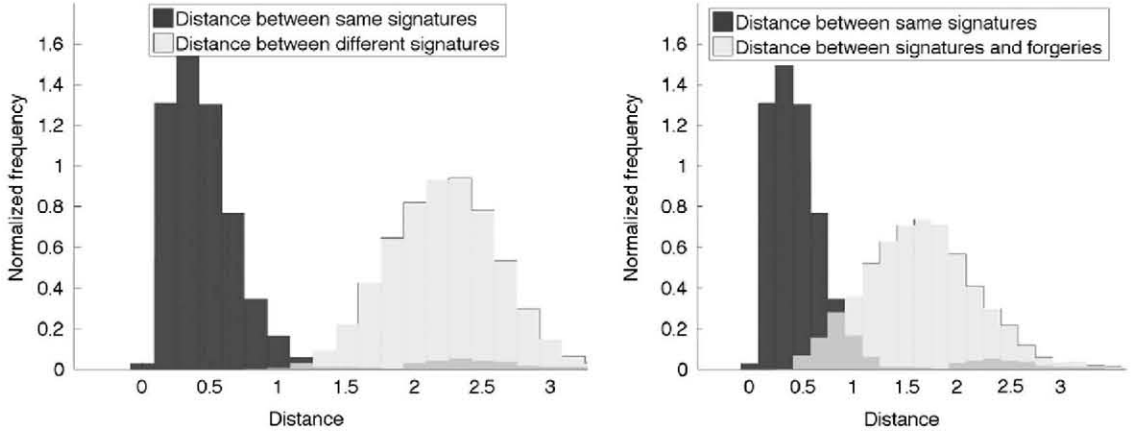

Fig. 10. Normalized histograms of distances between signatures and forgeries.

therefore signatures of different subjects are easily separable. Nevertheless, the distributions of distances between signatures of the same subject and their forgeries are overlapping, i.e. some forgeries are quite similar to the forged signatures.

In contrast to previously analysed methods, the DTW has no learning phase since it only stores the training instances. During the testing stage, this approach measures the similarity between the given instance and the training ones. In the case that there are k training instances for each signature, the most typical method for the classification process is k-nearest neighbour (k-NN). However, this assumes carrying out k comparisons between the given instance and that of training. In this paper, we propose another method called "Average DTW" which requires only one comparison. This method, which is based on the work presented in [36], generates a template from the N training instances and then compares this template to the analysed instance.

The first step of this algorithm is to calculate the DTW distances between all possible pairs of all training instances $I_{1..N}$ in order to decide which instance can be used as a reference. The reference instance I_R will be the one that presents the smallest average distance to all the remaining instances. We suppose that I_R is normalized and we normalize the remaining instances using their warping paths w_k in respect to this reference instance. Each warping path w^k indicates how to compress and expand the time scale of the instance I_k to minimize the distance with I_R. Between every pair of consecutive elements w_t^k and w_{t+1}^k of the warping path w^k, the DTW algorithm allows only three types of movement along the matrix M: horizontal, vertical and diagonal. Depending on these movements we have rescaled the instance I_k following the indications in Table 1.

Table 1
Pair-wise alignment criteria.

Movement	Action
Horizontal	Compressing samples $I_k(t)$ and $I_k(t+1)$
Vertical	Extending from sample $I_k(t)$ to $I_k(t+1)$
Diagonal	No rescaling

Compressing samples means joining them and this produces a loss in the original information. In order to minimize this loss, we combine these points by calculating their average. In the case of extending a sample, we must repeat the original value of the sample for the next samples. After rescaling all instances, for each instance we will obtain a normalized sequence which has the same length as the reference instance I_R. At this point, we propose to extract the average template of these instances by calculating the average points for each time sample. An example of the template extraction from several training instances is shown in Fig. 11.

Fig. 12 shows the evolution of EERs when increasing the number of training instances. The EERs for only one training instance (8.83% and 4.72%) correspond to the EERs obtained from the distributions depicted in Fig. 10 without any learning stage

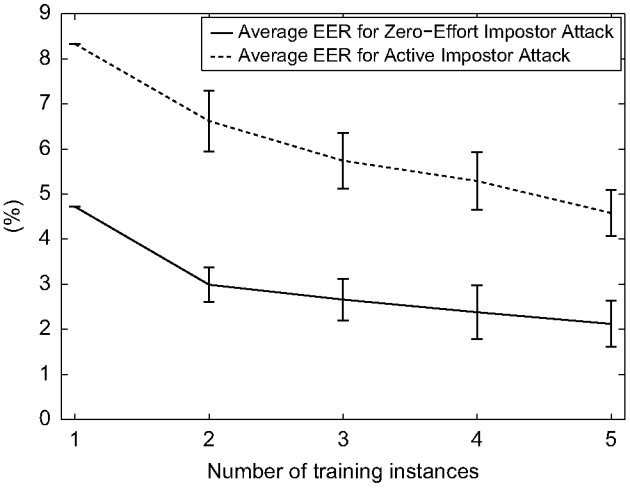

Fig. 12. Performance of average DTW for training sets with different sizes.

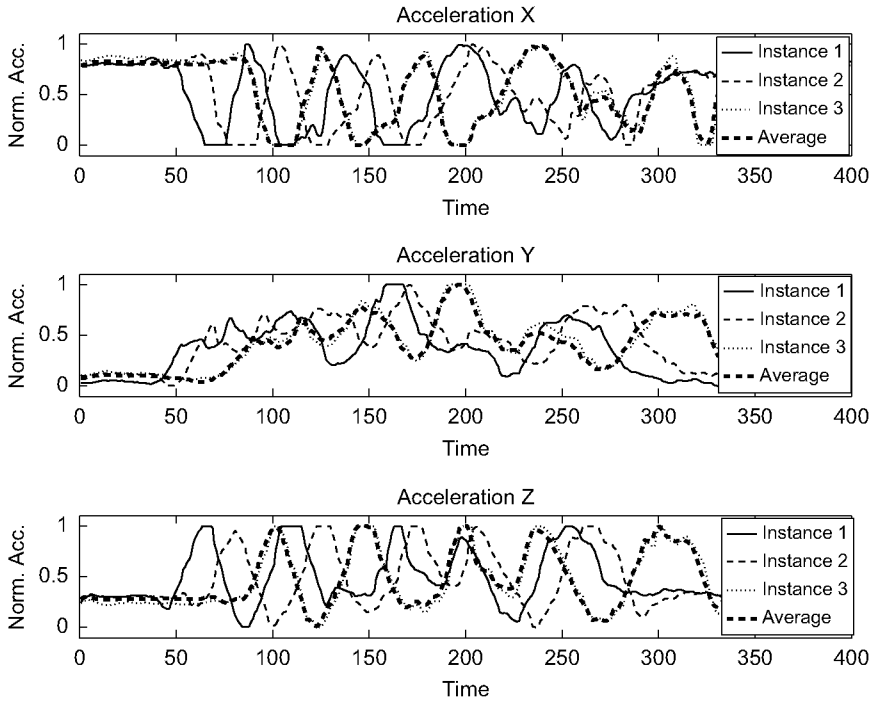

Fig. 11. Template extraction from three training instances using Average DTW.

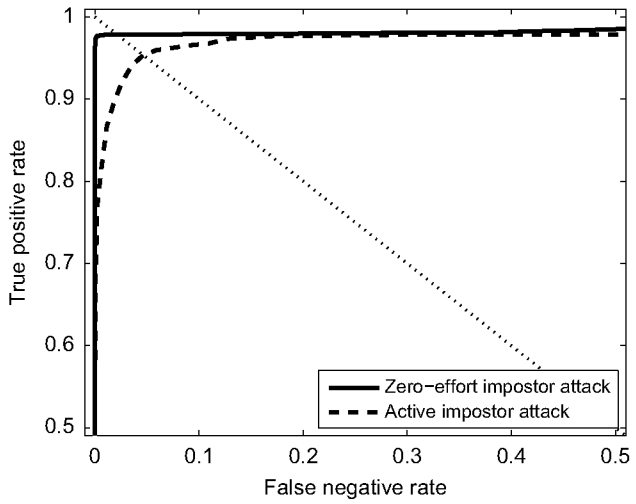

Fig. 13. ROC curves using an average template extracted from five training instances.

and they have been represented only to show the improvements achieved by the learning method. The best rates (4.58% and 2.12%) are reached by using five training instances however, should we need to reduce the enrolment time, we can also consider using only three training instances since their results are also acceptable (5.74% and 2.66%).

Finally, aiming to show the relation between FAR and FRR, in Fig. 13 we have represented the ROC curves for both attacks obtained by using five training instances. The ROC curve corresponding to the zero-effort impostor attack indicates that this technique produces an excellent classifier which provides high true positive rates and low false negative rates. On the other hand, the other ROC curve shows a decreasing accuracy of this technique when discriminating between signatures and their forgeries in the active impostor attack. Nevertheless, this performance decreasing is more attenuated than for previously analysed techniques. In order to show the differences between zero-effort attack and impostor attack, we have calculated the threshold values which produce EERs for both attacks obtaining 1.39 and 0.7, respectively. These values reflect that it is necessary to reduce the threshold by half to discriminate between signatures and forgeries.

6. Comparison of methods

In order to compare all the different methods, we have summarized the obtained results in Table 2. This table represents the EERs reached for each approach when using training sets of different sizes.

As expected, in all methods, the results for discriminating between forgeries and genuine signatures are rather worse than the results for distinguishing between the signatures of different individuals. Specifically, this difference is extremely high in the case of Bayesian classifiers since this method obtains the lowest EER when discriminating between individuals but it also reaches the highest EER for an active impostor attack. For this reason, we consider that the Bayesian classifiers are suitable for solving this problem though they may be fruitful in other fields such as gesture recognition.

In average, the best results are reached by the method based on DTW. For instance, the rates obtained by DTW without any kind of training are comparable to the rates obtained by HMM when using four training instances. However, the highest performance is achieved by means of the use of the training stage (average DTW) as the results obtained by using two training instances are comparable to the results of HMM with a training set of five instances.

As well as the analysis of recognition performance, we must also take into account the computational cost of each algorithm. All of the algorithms have been developed in Matlab from Mathworks Inc. and implemented on a computer with a Intel processor Duo T9400 and 4 GB of RAM. We have tested each algorithm with a signature sample with a length of 4.17 s which is the average length of all signatures. The execution time of DTW algorithm[1] to calculate the distance between the template and this sample is about 0.030 s whereas obtaining the forward probability with the HMM algorithm[2] took 0.042 s. The Bayesian classifier[3] could not be evaluated since the Matlab function, which generates this classifier, includes both training and testing stage in the same code, so we could not obtain the duration of the testing stage separately. However, in our opinion, this algorithm should have the lowest computational cost since it does not iterate over the temporal signal. Therefore, we have concluded that the pattern recognition technique which provides the best accuracy/computation cost relationship is average DTW.

7. Permanence analysis

Behavioural biometric techniques such as in-air signature must face the problem of repeatability. In our case, this means that a user has to be able to repeat the gesture with little variance. For this reason, we have carried out this permanence analysis which assesses the variability of in-air signature in time. For this analysis, we have used average DTW because this pattern recognition technique has provided the best performance in previous experiments. In this test, we generated a template for each signature using the five instances corresponding to the first day. Then, we calculated the distances between this template and the instances captured during the following days.

In Fig. 14, we have divided the evolution of these distances into two different groups of subjects. In the lower figure we have represented the users who are able to repeat their signature accurately over time because we can see that their distances remain stable. On the other hand, the upper figure contains increasing distances which represent those users who gradually change the performance of their signatures with the passing of the time. Although this means that the repeatability of in-air signature depends on the user's ability, we could deal with this variability by updating the template over time as the authors of [13] have proposed.

8. Conclusion and future work

In this work, we have shown the feasibility of using an in-air signature as a biometric technique for verification of identity. Although, the results are promising, this technique must be considered a weak biometric technique since the results are far from the results obtained by other well-known biometric techniques such as iris [37] or palmar surface of the hand [38]. Nevertheless, these results are comparable to other behavioural biometric techniques such as on-line handwritten signature [39] or gait recognition [40]. Hence, this technique must be relegated

[1] Implemented by Timothy Felty and available in http://www.mathworks.com/matlabcentral/fileexchange/6516-dynamic-time-warping.

[2] Implemented by Kevin Murphy and available in http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html.

[3] Implemented by Mathworks Inc.

Table 2
Average EERs for all studied methods using training sets of different sizes.

Method	Data	Number of training instances				
		1	2	3	4	5
HMM	Zero-effort	12.05% ± 0.75	7.98% ± 0.69	6.43% ± 0.32	5.45% ± 0.77	3.93% ± 0.96
	Active impostor	14.15% ± 0.70	10.17% ± 0.91	8.34% ± 0.61	7.47% ± 0.66	5.96% ± 0.55
Bayes	Zero-effort	N.A.	3.21% ± 0.73	2.28% ± 0.49	1.98% ± 0.51	1.81% ± 0.33
	Active impostor	N.A.	13.29% ± 0.87	13.93% ± 0.54	14.07% ± 0.56	14.09% ± 0.37
DTW	Zero-effort	4.72% ± 0	2.99% ± 0.39	2.66% ± 0.46	2.38% ± 0.6	2.12% ± 0.51
	Active impostor	8.83% ± 0	6.62% ± 0.68	5.74% ± 0.62	5.29% ± 0.64	4.58% ± 0.51

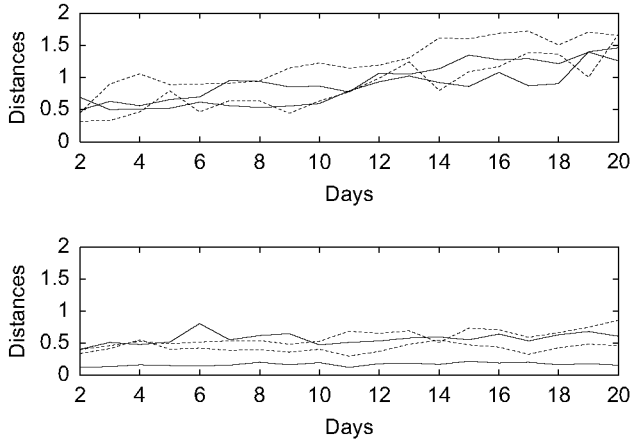

Fig. 14. Evolution in time of the distances between the initial template and daily captured instances.

to non-critical security applications or be combined with other techniques to reach the necessary accuracy.

After the comparison of different techniques for the recognition of in-air signatures, we can conclude that the most appropriate method to solve this problem is DTW. In particular, the "average DTW" algorithm which extracts a template from different training instances has obtained an EER of 2.12% when discriminating the signatures of several individuals. Furthermore, this technique has demonstrated its robustness against spoofing attacks since the EER only rose to 4.58% when six impostors attempted to forge the signatures after viewing recorded videos several times in which the individuals performed their in-air signatures.

Although, the results obtained by other techniques are inferior to the ones obtained by DTW, we consider that these techniques may be combined in order to improve their results. For instance, the Bayesian classifier presents the lowest EER 1.81% when discriminating between signatures of different individuals so this ability could be used in a prior stage of other technique to reject the signatures of other individuals. Furthermore, this technique present a low computational cost, thus it supposes that the combined technique will need less computational power and therefore it could be implemented in a mobile platform. A further study of the benefits of this combination will be carried out in future works.

References

[1] E.P. Lim, Mobile commerce: promises, challenges, and research agenda, Journal of Database Management 12 (2001) 4–13.
[2] G.N. Daldal, A cellular phone based home/office controller and alarm system, GU Journal of Science 19 (2006) 21–26.
[3] H. Hermanns, Mobile democracy: mobile phones as democratic tools, Politics 28 (2008) 74–82.
[4] A.K. Jain, P. Flynn, A.A. Ross, Handbook of Biometrics, Springer-Verlag, New York, Inc., Secaucus, NJ, USA, 2007.
[5] S.S. Chowhan, G.N. Shinde, Iris biometrics recognition application in security management, in: Congress on Image and Signal Processing, vol. 1, 2008, pp. 661–665.
[6] Y.J. Chin, T.S. Ong, M.K. Goh, B.Y. Hiew, Integrating palmprint and fingerprint for identity verification, in: International Conference on Network and System Security, 2009, pp. 437–442.
[7] V. Kanhangad, A. Kumar, D. Zhang, Combining 2d and 3d hand geometry features for biometric verification, Computer Vision and Pattern Recognition Workshop (2009) 39–44.
[8] C. Nandini, C.N. RaviKumar, Multibiometrics approach for facial recognition, in: International Conference on Computational Intelligence and Multimedia Applications, vol. 2, 2007, pp. 417–422.
[9] Y. Zhu, T. Tan, Y. Wang, Biometric personal identification based on handwriting, in: International Conference on Pattern Recognition, vol. 2, 2000, p. 2797.
[10] M. Rybnik, M. Tabedzki, K. Saeed, A keystroke dynamics based system for user identification, in: International Conference on Computer Information Systems and Industrial Management Applications, 2008, pp. 225–230.
[11] M.S. Nixon, T. Tan, R. Chellappa, Human Identification Based on Gait, Springer-Verlag, New York Inc., 2006.
[12] E. Farella, S. O'Modhrain, L. Benini, B. Riccó, Gesture signature for ambient intelligence applications: a feasibility study, Pervasive Computing (2006) 288–304.
[13] K. Matsuo, F. Okumura, M. Hashimoto, S. Sakazawa, Y. Hatori, Arm swing identification method with template update for long term stability, Advances in Biometrics (2007) 211–221.
[14] J.H. Steve Dowling, Nancy Paxton, Apple reports first quarter results, 2009.
[15] R.E. Haskell, D.M. Hanna, K.V. Sickle, 3d signature biometrics using curvature moments, in: International Conference on Artificial Intelligence, Las Vegas, Nevada, USA, pp. 718–721.
[16] J. Liu, L. Zhong, J. Wickramasuriya, V. Vasudevan, User evaluation of lightweight user authentication with a single tri-axis accelerometer, in: Proceedings of the 11th International Conference on Human–Computer Interaction with Mobile Devices and Services, ACM, p. 15.
[17] S. Kallio, J. Kela, J. Mantyjarvi, Online gesture recognition system for mobile interaction, in: IEEE International Conference on Systems Man and Cybernetics, vol. 3, 2003, pp. 2070–2076.
[18] A.K. Jain, F.D. Griess, S.D. Connell, E. Lansing, On-line signature verification, Pattern Recognition 35 (2002) 2002.
[19] J. Guo, D. Doermann, A. Rosenfeld, Local correspondence for detecting random forgeries, in: International Conference on Document Analysis and Recognition, 1997, p. 319.
[20] M. Kauppila, S. Pirttikangas, X. Su, J. Riekki, Accelerometer Based Gestural Control of Browser Applications, in: International Workshop on Real Field Identification (RFId2007). In conjunction with Fourth International Symposium on Ubiquitous Computing Systems (UCS 2007), pp. 25–28.
[21] J. Wu, G. Pan, D. Zhang, G. Qi, S. Li, Gesture recognition with a 3-D accelerometer, Ubiquitous Intelligence and Computing (2009) 25–38.
[22] C. Verplaetse, Inertial proprioceptive devices: self-motion-sensing toys and tools, IBM Systems Journal 35 (1996) 639–650.
[23] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE 77 (1989) 257–286.
[24] J. Hu, M.K. Brown, W. Turin, HMM based on-line handwriting recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (1996) 1039–1045.
[25] P. Lukowicz, J.A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, T. Starner, Recognizing workshop activity using body worn microphones and accelerometers, 2004.
[26] J. Mäntyjärvi, J. Kela, P. Korpipää, S. Kallio, Enabling fast and effortless customisation in accelerometer based gesture interaction, in: MUM '04: Proceedings of the 3rd International Conference on Mobile and Ubiquitous Multimedia, ACM Press, New York, NY, USA, 2004, pp. 25–31.
[27] T. Pylvänäinen, Accelerometer based gesture recognition using continuous HMMs, Pattern Recognition and Image Analysis 11 (2005) 639–646.

[28] V.M. Mäntylä, J. Mäntyjärvi, T. Seppänen, E. Tuulari, Hand gesture recognition of a mobile device user, in: Proceedings of the International IEEE Conference on Multimedia and Expo, pp. 281–284.
[29] J. Fierrez, J.O. Garcia, D. Ramos, J.G. Rodriguez, HMM-based on-line signature verification: feature extraction and signature modeling, Pattern Recognition Letters 28 (2007) 2325–2334.
[30] L. Bao, S.S. Intille, Activity recognition from user-annotated acceleration data, Pervasive Computing (2004) 1–17.
[31] I. Batal, M. Hauskrecht, A supervised time series feature extraction technique using DCT and DWT, in: Fourth International Conference on Machine Learning and Applications, 2009, pp. 735–739.
[32] T. Frantti, S. Kallio, Expert system for gesture recognition in terminal's user interface, Expert Systems with Applications 26 (2004) 189–202.
[33] Speech recognition using dynamic time warping with neural network trained templates, vol. 2, Baltimore, MD, USA, 1992.
[34] A. Corradini, Dynamic time warping for off-line recognition of a small gesture vocabulary, in: IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, IEEE Computer Society, Washington, DC, USA, 2001, pp. 82–89.
[35] R. Muscillo, S. Conforto, M. Schmid, P. Caselli, T. D'alessio, Classification of motor activities through derivative dynamic time warping applied on accelerometer data, in: EMBS 2007. 29th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, 2007, pp. 4930–4933.
[36] R. Muscillo, M. Schmid, S. Conforto, The median point DTW template to classify upper limb gestures at different speeds, in: Fourth European Conference of the International Federation for Medical and Biological Engineering, Springer, pp. 63–66.
[37] C. Sanchez-Avila, R. Sanchez-Reillo, Two different approaches for iris recognition using Gabor filters and multiscale zero-crossing representation, Pattern Recognition 38 (2005) 231–240.
[38] T. Savic, N. Pavesic, Personal recognition based on an image of the palmar surface of the hand, Pattern Recognition 40 (2007) 3152–3163.
[39] A.K. Jain, F.D. Griess, S.D. Connell, On-line signature verification, Pattern Recognition 35 (2002) 2963–2972.
[40] N. Boulgouris, K. Plataniotis, D. Hatzinakos, Gait recognition using linear time normalization, Pattern Recognition 39 (2006) 969–979.

Gonzalo Bailador (Madrid, Spain, 1980) received his M.Sc. in Computer Science from the Universidad Politécnica de Madrid, Spain, in 2003. Currently, he is pursuing a Ph.D. in Computer Science at the same university. His thesis is in the field of temporal pattern recognition. Since September 2009, he is working at CEDINT in the Biometric Department.